

A TARTALMI FELTÁRÁS KÉRDÉSEI 1945-TŐL AZ 1970-ES ÉVEKIG ÚTTÖRŐK ÉS ÚTKERESŐK

CSÍK TIBOR

TARTALMI ÖSSZEFOGLALÓ

A könyvtárak rendszerező és az ismereteket rendelkezésre bocsátó munkáját a XX. század közepén számos kritika érte. A szellemi és technológiai kihívásokra született egyik válasznak a számítógépes információ-visszakeresés kidolgozása, a másiknak pedig a könyvtári osztályozás, indexelés elméletének és gyakorlatának megújítása tekinthető az 1945-től az 1970-es évekig terjedő időszakban. A tanulmány célja egyrészt, hogy történeti áttekintést adjon az információ-visszakeresés megszületéséről és az osztályozás megújításáról, a benne részt vevő, tudós szakemberek munkássága alapján. A másik cél, azoknak a legfontosabb eszméknek, problémáknak és gyakorlati megoldásoknak a bemutatása, amelyek megalapozták a következő, az 1990-es évekig tartó korszak gondolkodását, és amelyek ma is hatnak.

Gernot Wersig (1936–2006) a német információtudomány nagy alakja az 1948-tól az 1970-es évekig tartó korszakot Shannon és Weaver fázisnak nevezi, amelyben a tudomány Shannon információ fogalmát igyekezett magához alakítani. A következő korszakot az online információszolgáltatások határozták meg, ez az időszak tehát egészen az 1990-es évekig tart.¹ Tanulmányunk célja, hogy áttekintést adjon arról, hogy a Shannon és Weaver fázis időszakában miként alakult a tartalmi feltárás elmélete és gyakorlata, kik voltak a korszak úttörői, útkeresői.

Szellemi és technológiai kihívások

A könyvtárak rendszerező és az ismereteket rendelkezésre bocsátó munkáját a XX. század közepén számos kritika érte. Az egyik legmeghatározóbb álláspontot Vannevar Bush (1890–1974), a tudós mérnök fejtette ki *As we may think*² című esszéjében, amely 1945-ben jelent meg. A másik kritikát a szellemi élet szépirodalomhoz kötődő területéről választottuk, Jorge Luis Borges (1899–1986) argentin írótól, gondolkodótól. Borges *El idioma analítico de John Wilkins* című esszéje³ az 1940-es években született, 1952-ben jelent meg kötetben, majd 1964-es angol fordításával vált általánosan ismertté.

Bush nemcsak a műszaki tudományok kiemelkedő alakja volt, hanem a második világháború alatt annak a hivatalnak a vezetője (Office of Scientific Research and

Development), amely összefogta a katonai fejlesztéseket az Egyesült Államokban. Így álláspontjának súlyt adott a tudományos elismertsége, de különösen politikai befolyása. Írásából három fontos megállapítást emelünk ki:

1. A tudományos publikációk nagy száma és eltérő minősége akadályozza a tudományos munkát, s a már meglévő eredmények fölhasználást. A kutatási eredmények átadásának, megismertetésének módszerei elavultak és a cél elérésére alkalmatlanok. Ez oda vezethet, hogy jelentős felfedezések elveszhetnek a tudomány számára. Intő példaként *Johann Georg Mendelnek*⁴ a genetikai törvényekről szóló munkáját hozza föl.
2. Az anyagok leválogatásának gépesítésében a könyvtárak lemaradtak, s hiányoznak az ehhez szükséges berendezések is. Az általuk alkalmazott indexelési, osztályozási eljárások mesterkélték, s a hierarchiában mindennek csak egy meghatározott helye van. Az eligazodáshoz ismerni és alkalmazni kell a szabályokat, majd a sikeres keresés után kilépni a rendszerből, s hogy új utat kezdjünk, megint belépni.
3. Az adatrögzítéshez és -továbbításhoz, valamint a leválogatáshoz azonban új eszközök állnak már rendelkezésre. Ezek egy része új műszaki fejlesztés, a többi már alkalmazzák az üzleti életben. A tárolásra, a másolatkészítésre és legfőképpen a szelekcióra jó megoldást kínál a mikrofilmalapú, fotomechanikus eljárás.⁵

Bush tehát nem tesz javaslatot a könyvtári feldolgozás megújítására. A megoldást a visszakeresés gépesítésében látja, ezért a fejlesztés középpontjába a technológiai kutatásokat állítja.

A műszaki haladást képviselő tudós után lássuk Borges, a könyvtáros, a szellem embere gondolatait. Az esszéjében Wilkins⁶ az univerzum rendszerét megalkotni kívánó „osztályozó” ember jelképe. Borges csak érzékelteti a Wilkins-féle osztályok, alosztályok szerveződésének elveit és azok jelölését, majd egy ősi kínai enciklopédia – *A jóra való ismeretek égi gyűjteménye* – osztályait sorolja föl:

„Amaz ősi feljegyzések szerint ugyanis az állatok lehetnek a) a Császár állatai, b) balzsamozottak, c) idomítottak, d) malacok, e) szirének, f) mesebeliek, g) kóbor kutyák, h) olyanok, amelyek ebben az osztályozásban szerepelnek, i) örülten rázkódók, j) megszámlálhatatlanok, k) olyanok, akiket a legfinomabb teveszőr ecsettel festettek, l) másfajták, m) olyanok, akik az imént törtek össze egy vázát, n) olyanok, akik távolról légynek tűnnek.

A Brüsszeli Bibliográfiai Intézet is csak gyarapítja a káoszt: ezer alcsoportra osztotta fel az univerzumot; a 262 a pápára vonatkozik; a 282 a római katolikus egyházra; a 263 az úrnapijára; a 268 a vasárnapi iskolákra; a 298 a mormonizmusra; a 294 pedig a brahmanizmusra, a buddhizmusra, a sintoizmusra és a taoizmusra. Heterogén alcsoportokat is megtűr, példaként lássuk a 179-et: „Állatokkal való kegyetlenkedés. Állatvédelem. A párbaj és az öngyilkosság az erkölcs szemszögéből. Különböző hibák és fogyatékoságok. Különböző erények és tulajdonságok.”

Bemutattam, hogy milyen önkényességek vannak Wilkins, az ismeretlen (apokrif) kínai enciklopédiaíró és a Brüsszeli Bibliográfiai Intézet munkájában; nyilvánvaló, hogy az univerzumnak nincs egyetlen osztályozása sem, amely ne volna önkényes és pontatlan.⁷

Borges bírálata közvetlenül az *Egyetemes Tizedes Osztályozás (ETO)* tartalmi és felosztásbeli hibáira vonatkozik. Következtetésében azonban megkérdőjelezi a *Paul Otlet* és *Henri La Fontaine*⁸ vezette dokumentációs mozgalom lelkesítő célját, hogy nemzetközi együttműködéssel számba vegyék és rendszerezék az emberiség tudását. Borges az ETO-t éppen olyan önkényesnek és hiábavalónak tekinti, mint John Wilkins „filozófiai nyelv”-ének a teremtett világot rendszerezve leíró szótárát vagy a képzeletbeli kínai enciklopédia taxonómiáját. Az esszé azt erősíti meg olvasójában, hogy a világról alkotott tudásunk és annak rendezése ahhoz a társalomhoz kötődik, csak abban érvényes, amely azt létrehozta.

A tartalmi feltárás eszközei a könyvtárban

A XX. század közepén – amikor a tudós és az író kifejtette álláspontját – a könyvtárakban a tartalmi feltárás általánosan elfogadott eljárásai az osztályozás és a tárgyszavazás voltak. Tekintsük át azokat a legfontosabb elveket és eszközöket, amelyeket a könyvtári gyakorlatban használtak a tartalmi reprezentáció elkészítésében, és amelyek alkalmazásával meg lehetett keresni egy téma irodalmát.

Melvil Dewey (1851–1931) 1876-ban megjelentetett elképzelései⁹ számos ún. tizedes osztályozásnak képezték az alapját. A két legjelentősebb, a későbbi elnevezés szerinti *Dewey Tizedes Osztályozás (DTO)* és a főntebb már említett ETO. Bár Dewey művében hivatkozik *Francis Bacon* (1561–1626) tudományrendszerezésére, osztályozása azonban nem filozófiai vagy tudományelméleti alapú, hanem korának, a XIX. század végének felsőoktatási gyakorlata volt a kiindulópontja. *Hjørland* – több meghatározó szakembert idézve – megállapítja, hogy a DTO elsődlegesen a könyvtári üzem racionális szabályozása szempontjából fontos, legnagyobb hatású elemének pedig az osztályozó ismérvek jelölésének rendszere tekinthető.¹⁰

Az ETO a DTO tudományfelosztásán némileg módosított, de a lényegét tekintve nem változtatott. A továbbfejlesztés fő irányai a polihierarchia átgondolt kiépítése és az alosztások rendszerének kidolgozása, valamint a jelzetek logikai összekapcsolhatóságának megteremtése voltak. Azt kívánták biztosítani, hogy a téma valamennyi szignifikáns elemét le lehessen írni. Ez gyökeres szakítást jelentett a monohierarchikus, enumeratív felfogással, és lehetőséget teremtett egy olyan osztályozásra, amelyben a témát több, egymástól független szempont szerint tárják föl. Bár az osztályozó ismérvek elrendezése a fő- és segéd táblázatokban, illetve a szakterületi alosztásokban is hierarchikus, annak minden előnyével (pl. rendezettség) és hátrányával (pl. eltérés a valóságtól, logikai következtelenség). De egyértelműen kiemeli a többi osztályozás

közül a téma elemző föltárását a rendszerben adott szempontok szerint, ami sok tekintetben előlegezi a fazettás osztályozásban kiteljesedő elveket.¹¹

A legáltalánosabban használt tárgyszójegyzék az Egyesült Államok nemzeti könyvtárának (Library of Congress – LC) tárgyszójegyzéke, a *Library of Congress Subject Headings (LCSH)*. A tárgyszavazás gyakorlata az Egyesült Államokban *Charles Ammi Cutter* (1837–1903) szótárkatalógus szabályzatára épült. Az LC az 1890-es évek végétől kezdte el építeni tárgyszójegyzékét, amelynek első kiadása 1914-ben jelent meg. A jegyzék kezdetben enumeratív volt, és két relációt használt: a *lásd* (see) és a *lásd még* (see also). A folyamatos fejlesztések révén – például tematikus, formai, földrajzi, idő alosztások alkalmazása – majd a tárgyszavazási szabályzat kidolgozásával¹² az 1950-es évekre egyre jobban előtérbe került a tartalom elemző megközelítése.

Az LC osztályozási rendszerének (*Library of Congress Classification – LCC*)¹³ kidolgozása is az 1890-es évek legvégén kezdődött. Az elsődleges cél meglehetősen gyakorlati volt: a könyvtár egyre növekvő gyűjteményének megfelelő föltárása oly módon, ami biztosítja a könyvtár szervezetének hatékony működését és az állomány jó fizikai elrendezhetőségét. A különböző tudományágak, szakterületek táblázatai az 1900-as évek elejétől egészen a század közepéig folyamatosan jelentek meg. A már kidolgozott táblázatokat a munka közben fölmerült szempontok alapján újra és újra átdolgozták, bővítették, aminek következtében a gyűjteményrészeket újra kellett szakszerezni, akár többször is. Az LCC-t az enumeratív osztályozási rendszerek példaként szokás említeni, bár egyes tudásterületek osztályában használ alosztásokat. Az LCC jellemzője, hogy a különböző tudásterületek táblázatai eltérő elvek szerint épültek föl, ami az adott területért felelős szervezeti egység önállóságából is következett. Az osztályozási rendszer egészét átfogó elvek hiánya mellett általánosan jellemző az osztályozó ismérvek monohierarchikus elrendezése. A hierarchia azonban nem követi az alárendelés (szubordináció) szigorú logikai szabályait, sokszor tematikus csoportosítást, illetve aláosztást (szubszumció) valósít meg.

Henry Evelyn Bliss (1870–1955) alkotta meg a XX. század közepének legelismertebb osztályozását, az ún. *Bibliographic Classification (BC)*¹⁴, amely 1940–1953 között jelent meg. Bliss áttanulmányozta, elemezte kora osztályozási rendszereit, s a levont következtetések alapján dolgozta ki az új osztályozást. De nemcsak gyakorlati szakember volt, hanem az osztályozáselmélet úttörője, és az ő nevéhez fűződik az ismeretszervezés (knowledge organization) fogalmának megalkotása, tartalmának meghatározása is. Munkásságából kiemelendő az a gondolat, hogy a könyvtári osztályozásnak nemcsak koherensnek, átfogónak, logikai elveken alapulónak kell lennie, hanem összhangban kell állnia a tudomány rendszereivel. Ahogy a tudomány tükrözi a természetet, akképpen kell tükröznie a tudományt a könyvtári osztályozásnak.

Bliss elmélete óriási hatást gyakorolt *Siyali Ramamrita Ranganathan*-ra (1892–1972) a XX. század leginvenciózusabb osztályozási szakemberére, a könyvtártudomány egyik megalapozójára. Ranganathan *Kettőspontos Osztályozásában (Colon*

Classification – CC) egyesítette a nyugati osztályozásméletet és a hindu világszemléletet, továbbá követelménnyé emelte az osztályozás folyamatában az analízist és a szintézist. Osztályozása a fogalmi kategóriák elméletéhez áll közel, amely megkülönbözteti az entitást (Arisztotelésznél ουσία / szubsztancia) és annak attribútumait. A téma föltárása ún. fazetták szerint történik, amelyek a következők lehetnek: *Personality* (egyesediség, a téma fókusza), *Matter* (tulajdonság, alkotóelem), *Energy* (folyamat, tevékenység, hatás), *Space* (térbeli, földrajzi vonatkozás), *Time* (időbeli vonatkozás). Egy osztályozó ismerv kerülhet a P fazettába, de értelemszerűen bármely más fazettába is, attól függően, hogy az adott téma melyik vetületét írja le. Ranganathan fazettás osztályozásának újdonságai közül két szempontot ki kell emelni: az egyik, az osztályozó ismérvek csoportosítása az adott szakterület legáltalánosabb fogalmai szerint, a másik, az osztályozó ismervnek az adott témában betöltött szerepének jelölése.

Összefoglalva megállapítható, hogy általános jelenség az enumertív, monohierarchikus rendszertől való távolodás. Egyre fontosabbá válik, hogy egy téma leírását több osztályozó ismerv együttese alkossa. Több ismerv használata azonban fölveti a kérdést, hogy az egyes ismérvek mely tartalmi szempontot, a téma mely vetületét írják le. A tartalmi reprezentáción belül az ismérvek egymáshoz való viszonyának vizsgálata a kutatás új távlatait nyitotta meg. A föltárás során a téma elemzését és szabályozott leírását valamennyi ismertetett eszköz elvárja, bár ennek mértéke és módja óriási különbségeket mutat. Az élő tudomány és a könyvtári osztályozás viszonya is lényeges kérdés: a könyvtári rendszernek követnie kell-e a tudományok rendszertanát vagy más, például a gyűjtemény vagy a használók gyakorlatias elvárásainak kell megfelelnie. Mindez csak aláhúzza az osztályozásméleti kutatások fontosságát, a tartalmi reprezentáció tudományos kutatásának nélkülözhetetlenségét.

Az információ-visszakeresés

A második világháború után a természettudományos, műszaki szakirodalomnak és a kutatási dokumentációnak a feldolgozása és visszakereshetővé tétele a katonai vonatkozások miatt elvált a „polgári” fejlesztésektől. Az Egyesült Államokban Bush intenciói szerint az információs technológia, elsősorban pedig a leválogatás gépesítése állt a középpontban. A kutatási tevékenységek és a műszaki fejlesztések kormányzati támogatása 1950-re intézményesült. Ebben az évben – Bush hathatós közreműködésével – létrejött a *National Science Foundation (NSF)* az alap kutatások és a mérnöki tudományok segítésére. A következő lépést az 1958-as ún. Szputnyik-törvény (*National Defense Education Act – NDEA*) jelentette, amely az Alap feladatává tette az információs kutatások szervezését, nevesítve az indexelés, a kivonatkészítés, a fordítás, a tudományos információk disszeminálásának, elérhetővé tételének gépesítését.¹⁵

Az információellátás fontosságára ráirányította a figyelmet a Szovjetunió központi információs intézetének hatékony működése is, amely döntően járult hozzá a sikeres

katonai fejlesztésekhez. Az intézet a tudományos akadémia részeként jött létre 1952-ben, 1955-től viselte az Össz-szövetségi Tudományos és Műszaki Információs Intézet (*Всесоюзный институт научной и технической информации – VINITI*) nevet. A VINITI az egész világról, de különösen a nyugati országokból gyűjtötte és dolgozta föl a legfrissebb tudományos és műszaki közleményeket, s látta el információval a szovjet fejlesztéseket. A tudományos tájékoztatás magas színvonala, szervezettsége elsősorban a kivonatokat készítő, a dokumentációs tevékenységet végző tudósoknak, szakembereknek volt köszönhető, akik nem mindig „önszántukból” vettek részt a munkában. A VINITI nemcsak az információellátást biztosította kiadványaival és szolgáltatásaival, hanem összefogta az információkezeléssel kapcsolatos kutatásokat is.¹⁶

Az amerikai kutatások első, átütő eredményét *Claude Elwood Shannon* (1916–2001) érte el, aki Bush tanítványa volt és munkatársa a fotomechanikus leválogató fejlesztésében. Az 1949-ben megjelent műve lefektette a matematikai információelmélet alapjait, mérhetővé tette a kommunikációs rendszerek hatékonyságát.¹⁷ A hatása óriási volt, a tudományok széles köre vette át fogalmait, adaptálta modelljeit. A kommunikációs modellt (adó-csatorna-vevő) például évtizedeken át alkalmazták a könyvtárak tájékoztató tevékenységének bemutatására is. Pedig Shannon matematikai elmélete nem kapcsolódott az információkereséshez, sem a tudásreprezentációhoz.

Az információ-visszakeresés (information retrieval) fogalmát 1950-ben *Calvin Northrup Mooers* (1919–1994) határozta meg. A matematikus Mooers a téma szerinti visszakeresés gépesítését azzal a frusztrációval indokolta, amelyet a könyvtárak – kivétel nélkül valamennyi – ilyen irányú működése vált ki.¹⁸

Mortimer Taube (1910–1965) dolgozott a washingtoni Kongresszusi Könyvtárban, tanított a felsőoktatásban és vezette *Oak Ridge National Laboratory* dokumentációs rendszerének kialakítását. A nevéhez fűződik a dokumentum szövegéből kivont szavakkal való indexelés (Uniterm) és az erre alapozott gépi keresés kidolgozása, ami valódi mérföldkőnek számít.¹⁹ Bár a gépe még lyukkártyákkal működik, de a kulcsszóindex és a visszakeresésben a kulcsszavak Boole-operátorokkal való „koordinálása” ma is alkalmazott eljárás.

Hans Peter Luhn (1896–1964) az információs technológia legnagyobb újtója. Ő is lyukkártyás gép fejlesztésével kezdte, ahogy Mooers és Taube. Majd olyan gépi indexelési eljárásokat dolgozott ki, amelyek a közlemények címéből, szövegéből vonták ki a témára utaló szavakat, mint például a *KWIC*.²⁰ Lefektette továbbá az automatikus kivonatkészítés alapjait, megfogalmazva, hogy a szövegben szerepelő szavak előfordulási gyakorisága és a mondaton belüli relatív helyzete alapján meghatározható és mérhető a mondat szignifikanciája.²¹ Az ő nevéhez fűződik továbbá a szelektív információterjesztés (Selective Information Dissemination) elve, amelynek mai megfelelője az RSS.

Az 1950-es évek végén született meg az információkeresés új eszköze a tezaurus.²² A szót visszakereséssel kapcsolatban Mooers használta először, a mai értelmezéshez legközelebb Luhn felfogása áll, az elméletének megalapozója pedig *Brian Campbell Vickery* (1928–2009) volt.²³ A tezaurus az ismérvek megnevezését szabályo-

zottan tartalmazza (homonímia kezelése, szinonimák megadása); jelöli az ismérvek közötti logikai kapcsolatokat (pl. nem-faj viszony) és megadja a fogalmi szinteket, a tematikai összetartozást; az ismérvek között megadott relációk a gépi keresésben felhasználhatók (pl. az alárendelt ismérvek keresése); ugyanakkor indexnyelvként posztkoordinált; az osztályozási rendszerek felosztási kötöttségei nem érvényesek.²⁴ Először a DuPont vegyipari cég használt tezaurszt 1959-ben, 1960-ben jelent meg a *Thesaurus of Armed Services Technical Information Agency (ASTIA)*, majd a *Chemical Engineering Thesaurus* az American Institute of Chemical Engineers kiadásában. Mindez jól mutatja, hogy a tezaurszok elsődleges felhasználója az ipari és katonai terület volt.

Eugene Garfield (1925–2017) a tartalmi feltárás gépesíthetőségét kutatva vizsgálta a tudományos közlemények hivatkozásait.²⁵ Az általa kifejlesztett hivatkozási index azonban az információkeresésnek egy teljesen új dimenzióját nyitotta meg. A hivatkozások száma és különösen kapcsolatai (pl. együttidézés, hivatkozási háló) vizsgálata nem a tematikus visszakeresést, hanem a bibliometriát újította meg teljesen. Sok évtizedes munkásságának eredményei alapvetőek a tudományos teljesítmény mérésében, a tudományteréképezésben. Munkásságának korszakos jelentőségét az is bizonyítja, hogy számos hivatkozásvizsgáló eljárása előlegezi a webes linkelemző algoritmusokat.

Gerard Salton (1927–1995) tudományos pályája az 1950-es évek végén indult, de a 60-as évektől haláláig folyamatosan jelentkezett munkáival.²⁶ Az általa kidolgozott vektortér modell alapján megszülettek azok a számítógépes algoritmusok, amelyek meg tudják határozni a szövegek tárgyát és képesek azokat tematikus hasonlóságuk alapján rendezni. Kifejlesztett továbbá egy *SMART (System for the Mechanical Analysis and Retrieval of Text)* nevű információfeldolgozó és -visszakereső rendszert is.²⁷ Az automatikus osztályozásnak nevezett eljárása nem a hagyományos, logikai alapú intellektuális művelet. Az osztályba sorolás (klaszterálás) ugyanis nem egy vagy néhány lényeges ismértv alapján történik, hanem valamennyi számba vett ismértv figyelembevételével, sőt, az egyes ismérvek súlyozhatók is.

A vektortér modell mellett a visszakeresés ún. valószínűségi modellje volt a legnagyobb hatású. Az 1960-as évektől folyó kutatások egyik legkiemelkedőbb alakja Karen Ida Boalch Spärck Jones (1935–2007) volt. Visszanyúlva Luhn azon elgondolásához, amelyben a szavak előfordulási gyakoriságára alapozódik azok szignifikanciája, az egyediséget jelölő szavak előfordulását vizsgálta. A legtöbbet idézett matematikai eljárása az ún. inverz dokumentumgyakoriság (inverse document frequency), amely egy szó egyediségét, specifikusságát a dokumentumban való előfordulási gyakoriságának inverze alapján súlyozza és alkalmazza a visszakeresésben.²⁸ Munkássága az 1970-es években bontakozik ki, és ma is megkerülhetetlen a gépi információkeresésben, s erős hatása alól nem vonhatta ki magát a könyvtártudomány sem.

A Bush által kitűzött célok tehát teljesültek, a téma szerinti leválogatás, vagy ahogy most már általánosan nevezik, az információ-visszakeresés gépesítése megtörtént. Az eddig kizárólagosan emberinek tartott intellektuális műveletet, az osztályozást és a szig-

nifikancia szerinti sorba rendezést matematikai modellek alapján számítógépek végzik. A megszületett eljárások – szövegstatistikai módszer, vektortér modell, valószínűségi modell, hivatkozásmintázatok – egészen az 1990-es évekig meghatározóak.

A könyvtári osztályozás fejlesztése – útkeresés a tartalmi feltárás területén

Az osztályozásfejlesztés legfontosabb műhelyének az angol *Classification Research Group* (CRG) tekinthető, amely a Royal Society kezdeményezésére jött létre 1952-ben. A CRG 1955-ben megjelentetett manifesztuma a fazettás osztályozást határozza meg mint az információ-visszakeresés alapját.²⁹ A munkát nagyban segítette, hogy a NATO, osztályozás kialakítása címén, támogatást adott a csoportnak.³⁰

A CRG tevékenységéhez kapcsolódóan *Brian Campbell Vickery* (1918–2009) kidolgozta a fazettás osztályozás általános elméletét.³¹ Munkássága átfogta az információtudomány szinte valamennyi lényeges területét. Az információfeldolgozás intellektuális és gépi megközelítésének integrációjára törekedett, művei máig vonatkoztatási pontnak számítanak.

Janson Farradane (1906–1989) az információtudomány névadója, 1963-ban ilyen címmel hirdette meg egyetemi kurzusát (*City University, London*).³² Az új megnevezés jelezte, hogy mind a könyvtártudományt, mind a dokumentációt szükséges újradefiniálni. Oktatói, kutatói, tudományszervező és szakmai közéletet támogató tevékenysége (pl. az *Institute of Information Scientist*, a *Chartered Institute of Library and Information Professionals* elődszervezetének kezdeményezése) jól mutatja, hogy az újradefiniálás fontos mozzanata a „hagyomány” és az információs technológia szintézise. Munkásságának fontos része volt egy olyan indexelési rendszer kidolgozása, amelyben az ismérvek közötti relációt operátorokkal lehet jelölni, ezáltal is pontosítva a témát a gépi keresésben.³³

A tartalmi ismérvek témán belüli (szintaktikai) relációinak leírása az egyike volt a legfontosabb kutatási területnek. Az indexelés szintaktikai kérdéseivel foglalkozók közül Gardin és Austin munkásságát emeljük ki. *Jean-Claude Gardin* (1925–2013) vezette a francia nemzeti kutatási központ (*Centre national de la recherche scientifique – CNRS*) automatikus dokumentációt kutató szekcióját is. Az 1960-as évek elejére az EURATOM támogatásával kifejlesztett egy *Syntagmatic Organization Language* (SYNTOL) nevű indexelési eljárást, amely sok tekintetben megújította a dokumentációs nyelv elméletét.³⁴ *Derek William Austin* (1921–2001) tevékenyen részt vett a CRG munkájában, de legismertebb alkotása a British Library számára kidolgozott indexelési eljárás, amelynek neve *Preserved Context Index System* (PRECIS).³⁵ A PRECIS célja, hogy a tartalmi ismérveknek a dokumentum témájában meglevő viszonyát olyan módon írja le, hogy azt a gépi visszakeresés során hasznosítani lehessen.

Az osztályozási, indexelési rendszerek hatékonyságát vizsgáló kutatások közül a *Cyril William Cleverdon* (1914–1997) vezetésével lezajlottaknak volt a legnagyobb

visszhangja. Cleverdon bár angol volt, nem tartozott a CRG-hez, a kutatásait az amerikai NSF támogatta. Az ún. első Cranfieldi jelentés³⁶ négy rendszer – az ETO, a tárgyszavazás, a fazettás osztályozás és a UNITERM – hatékonyságát vizsgálta a visszakeresésben. Ha kis különbséggel is, a UNITERM teljesített legjobban, míg a fazettás osztályozás a legkevesbé jól. Bár a Cranfieldi jelentéseket nagyon sokan kritizálták, mégis korszakos jelentőségük volt a rendszerközpontú megközelítés és a rendszer értékelésének megalapozása, a relevancia, a teljesség és a pontosság gyakorlati alkalmazásával. (Bár Taube a relevanciát pszeudomatematikai mutatónak tartotta.)

A korábban tárgyalt osztályozási rendszerek fejlesztése is folyamatos volt. Az ETO megújítása az 1960-as, 70-es években gyorsult föl, s vált az osztályozási rendszer talán legsikeresebb időszakává. A nemzetközi együttműködéssel folyó munkálatok során modernizálták a tartalmat, hogy minél jobban igazodjon az élő tudomány rendszeréhez. Az osztályozásban fokozottan érvényesítették az analitikus-szintetikus jellegét, minél többet építve be a fazettás elvekből. Bliss osztályozásának (BC) továbbfejlesztésére 1967-ben társaság alakult, amely a CRG deklarálta fazettás, elemző elvek szerint dolgozta át a rendszert.³⁷

Ingetraut Dahlberg (1927–2017) mind az elmélet, mind a gyakorlat oldaláról jól ismerte az osztályozás problémáit, számos osztályozási rendszer fejlesztésében vett részt³⁸, többek között tagja volt a FID ETO fejlesztésével foglalkozó bizottságának is (*Revision Committee for the Universal Decimal Classification*). Álláspontja szerint az osztályozásnak fogalmi szinten kell megvalósulnia, ezért az osztályozási rendszernek egyetemes elvek szerint kell fölépülnie, s a rendszerezésnek filozófiai, tudományelméleti alapúnak kell lennie.³⁹ Dahlberg megújította és továbbfejlesztette az osztályozás klasszikus elméletét (építve pl. az arisztotelészi kategóriákra), ezzel és a tudomány-szervező tevékenységével megerősítette a nem gépi, hanem az intellektuális tartalmi feltárás helyzetét a tudományos diskurzusokban.

Jóllehet az információ-visszakereséshez kapcsolódóan vetődött föl a tartalmi reprezentáció új eszköze a teaurusz, de a könyvtári szakemberek gyorsan adaptálták az új „technológiát”. A teauruszok fejlesztésének óriási támogatást biztosított a bibliográfiai adatbázisok kiépítése. Az ipari fejlesztésekhez kapcsolódó tudományok teauruszai mellett sorra jelentek meg más tudományterületek teauruszai is. Például az *Education Research Information Center (ERIC)*, az Egyesült Államok pedagógiai adatbázisának teaurusza 1966-ra készült el. A gépi visszakeresés természetesen a többi ellenőrzött szótár fejlesztését is támogatta, de a teauruszok kiemelkedő jelentőségét az 1970-es években meginduló szabványosítás is bizonyítja.⁴⁰ A szakirodalmi adatbázisokhoz kidolgozott teauruszok, speciális osztályozások jelentősen erősítették a tartalmi reprezentáció domén alapú megközelítését is.

Az áttekintésből jól látható, hogy a könyvtári osztályozások fejlesztésének igen fontos iránya volt a hierarchia és a fazettás elrendezés megfelelő egyensúlyának kialakítása. Általánosan elfogadottá vált azon elv, amely a reprezentációt a téma valameny-

nyi lényeges ismérve szintézisének tekinti. A fejlesztések során arra törekednek, hogy az osztályozási rendszereket minél jobban közelítsék az élő tudomány szerkezetéhez. Elsősorban a fazettás elvekből kiindulva megszületett az az indexelési eljárás, amely reprezentálni kívánta az ismérveknek a dokumentumban meglévő viszonyait, mintegy indexelési szintaxis (PRECIS). A korszak útkeresésének része az emberi tudást mint egészet egységes elvek szerint elrendező osztályozási rendszerek kialakítása (vö. Dahlberg, BC2). De egyre jelentősebbé vált az egyes szakterületek tartalmi reprezentációját biztosító teauruszok, speciális osztályozások fejlesztése is, ami elsődlegesen a bibliográfiai adatbázisok kiépítéséhez kapcsolódott.

Összegzés helyett

A Shannon és Weaver fázis legnagyobb hozadéka, hogy a tartalmi reprezentáció valamennyi kérdését fölvetette. A tartalmi reprezentáció fejlődési ívét megrajzolva a kiindulás a lényegi ismérv alapján történő besorolás lenne, amely után a több szempont, vetület együttes figyelembevételével történő föltárások következnek, majd pedig a számítógépes, matematikai modell alapú, akár valamennyi jellemzőt – egyeseket súlyozottan is – számba vevő eljárások. A korszakban többen is keresik arra a kérdésre a választ, hogy melyek azok a rendszerező elvek, amelyek az emberi tudás egészében érvényesek, amelyek alapján szervezhetők az ismeretek. Vagy nem átfogó elveket kell keresni, hanem az egyes tudásterületek, ismeretkörök szerkezetére alapozni a tartalmi reprezentációt. Az automatikus osztályozás és referálás, az idézettségi mintázatok és rangsorolás – sokak szerint – elháríthatják Borges vádját az önkényességről és pontatlanságról, hiszen a rendezés a szövegből kiindulva matematikai modellek alapján történik. De ők még nem ismerhették a PageRank algoritmusokat...

A könyvtár- és információtudomány szakemberei nincsenek egyedül, amikor a rendszerezhetőség elveit kutatják. A természettudományokban például a rendszerezés (taxonómia) mindig is alapvető tudományos probléma volt. A könyvtári osztályozás elvei és azok változásai párhuzamba állíthatók az élőlények tudományos osztályozásának történetével – hívja föl a figyelmet Hjørland,⁴¹ – korunk információtudományának kiemelkedő alakja – Brent Mishler korszakolására hivatkozva.

Brent Mishler a biológia történetét vizsgálva a következő rendszertani korszakokat különíti el⁴²:

1. Prehisztórikus kor – „folk classification”.
2. Az ókori görögtől Linné taxonómiájáig terjedő időszak – az osztályozás alapelve az ún. esszencializmus. A hierarchikus rendszerezés szigorú logikai elvek szerint, a leglényegesebb ismertetőjegy alapján történik. A korszak két meghatározó alakja *Arisztotelész* (Kr. e. 384 – Kr. e. 322) és *Carl von Linné* (1707–1778) voltak.

3. A természet rendszere – a legszélesebb értelemben vett hasonlóság elvére épít, amelynek megállapításához több, fontos tulajdonságot is figyelembe vesz. Az *Antoine-Laurent de Jussieu* (1748–1836) kidolgozta módszer hosszú ideig általánosan elfogadottnak számított.
 4. *Charles R. Darwin* (1809–1882) evolúcióelmélete nem volt a rendszertanra alakító hatással egészen az 1950-es évekig – lásd a filogenetikus rendszertan.
 5. Numerikus taxonómia (fenetika) – matematikai (számítástechnikai) alapozású módszertan. Az 1950-es évek végétől, a 60-as évek elejétől alkalmazzák. Az volt az alapelve, hogy lehetőség szerint valamennyi tulajdonság figyelembevételével állapítandó meg a hasonlóság és hozandók létre az osztályok (cluster).
 6. Filogenetikus rendszertan (kladisztika) – az evolúciós információt hordozó ismérvek a taxonómia alapja. Az 1970-es években elismertté váló módszer *E. H. Willi Henning* (1913–1976) leszármazási mintázatokra összpontosító elveit követi.
 7. Molekuláris filogenetika – a DNS-vizsgálatok módszereire és eredményeire támaszkodva az 1980-as évek végétől hozott új rendszertani megközelítést.
- Mishler szerint alapvető változást, paradigmaváltást az esszencializmus, majd a több, fontos tulajdonságok szerinti elrendezés, s ezután a filogenetikai alapú taxonómia hozott.

Irodalom és jegyzetek

1. WERSIG, Gernot: Information theory. In: Feather, J. – Sturges, P. (eds.) International encyclopedia of library and information science. London, New York, Routledge, 2003. 313. p.
2. BUSH, Vannevar: As we may think. = Atlantic Monthly, 7. 1945. Forrás: <http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/3881> [2019. július 10.]
3. BORGES, Jorge Luis: El idioma analítico de John Wilkins. In: Uő.: Otras Inquisiciones (1937–1952). Buenos Aires, Sur, 1952.
4. Mendel-szindróma: az 1866-ban publikált genetikai szabály visszhangtalan maradt, 1900-ban újból felfedezte három tudós, Hugo de Vries, Carl Correns és Erich von Tschermak.
5. Emanuel Goldberg, a német Zeiss vállalat mérnökeként szabadalmaztatott mikrofilmes keresője volt a kiindulása Bush és tanítványai kifejlesztette szelektornak.; BUCKLAND, Michel K.: Emanuel Goldberg, electronic document retrieval, and Vannevar Bush's Memex. = Journal of the American Society for Information Science, 43. vol. 4. no. 1993. 284–294. p. Forrás: <http://people.ischool.berkeley.edu/~buckland/goldebush.html> [2019. július 10.]
6. A címben szereplő Wilkins valós személy, az angol Royal Society megbízásából a tudományok műveléséhez használható nyelvet dolgozott ki – WILKINS, John: An essay toward a real character and a philosophical language. London, 1668. Forrás: <https://archive.org/details/AnEssayTowardsARealCharacterAndAPhilosophicalLanguage/page/n37> [2019. július 10.] A filozófiai nyelvről bővebben MÉDER Zsombor-Zoltán: A „megfelelő”

- szavak és a dolgok – John Wilkins analitikus nyelve. = Kellék, 27–28. sz. 2005. Forrás: <http://kellek.adatbank.transindex.ro/pdf/27-28/033meder.pdf> [2019. július 10.]
7. BORGES, Jorge Luis: A John Wilkins-féle analitikus nyelv. In: Uő.: Az örökkévalóság története. Esszék. Ford. Boglár Lajos, Ertl István, Scholz László. Budapest, Európa, 1999. 276–285. p. <ftp://ontologia.hu/Language/Hungarian/Crawl/MEK/mek.oszk.hu/00400/00461/html/borges23.htm#xxvii> [2019. július 10.] A szövegrész angolul: „In its remote pages it is written that the animals are divided into: (a) belonging to the emperor, (b) embalmed, (c) tame, (d) sucking pigs, (e) sirens, (f) fabulous, (g) stray dogs, (h) included in the present classification, (i) frenzied, (j) innumerable, (k) drawn with a very fine camelhair brush, (l) et cetera, (m) having just broken the water pitcher, (n) that from a long way off look like flies. The Bibliographic Institute of Brussels exerts chaos too: it has divided the universe into 1000 subdivisions, from which number 262 is the pope; number 282, the Roman Catholic Church; 263, the Day of the Lord; 268 Sunday schools; 298, mormonism; and number 294, brahmanism, buddhism, shintoism and taoism. It doesn't reject heterogene subdivisions as, for example, 179: „Cruelty towards animals. Animals protection. Duel and suicide seen through moral values. Various vices and disadvantages. Advantages and various qualities.” I have registered the arbitrariness of Wilkins, of the unknown (or false) Chinese encyclopaedia writer and of the Bibliographic Institute of Brussels; it is clear that there is no classification of the Universe not being arbitrary and full of conjectures.” Borges, Jorge Luis: *The analytical language of John Wilkins*. Forrás: <https://ccrma.stanford.edu/courses/155/assignment/ex1/Borges.pdf> [2019. július 10.]
 8. Otlet (1868–1944) és La Fontaine (1854–1943) 1895-ben alapította az Office International de Bibliographie (IIB) nevű szervezetet, amelynek neve 1937–1988 között Fédération Internationale de Documentation (FID) volt.
 9. DEWEY, Melvil: A Classification and Subject Index for Cataloguing and Arranging the Books and Pamphlets of a Library. Amherst, 1876. Forrás: <http://www.gutenberg.org/files/12513/12513-h/12513-h.htm> [2019. július 10.]
 10. HJØRLAND, Birger: What is knowledge organization (KO). = Knowledge Organization, 35. vol. 2–3. no. 2008. 88–89. p. Forrás: <https://doi.org/10.5771/0943-7444-2008-2-3-86> [2019. július 10.] A bírálatok közül álljon itt példaként egy Bernd Frohmantól vett idézet: „Dewey disdained any philosophical excogitation of the meaning of his class symbols, leaving the job of finding verbal equivalents to others. His innovation and the essence of the system lay in the notation. The DDC is a poorly semiotic system of expanding nests of ten digits, lacking any referent beyond itself...” Uo. 89. p.
 11. HJØRLAND, Birger: Facet analysis. The logical approach to knowledge organization. = Information Processing & Management, 49. vol. 2. no. 2013. 545. p. Forrás: <https://doi.org/10.1016/j.ipm.2012.10.001>, https://www.academia.edu/3720713/Facet_analysis_The_logical_approach_to_knowledge_organization [2019. július 10.]
 12. HAYKIN, David Judson: Subject headings. A practical guide. Washington, Library of Congress, U.S. Government Printing Office. 1951. Forrás: <https://babel.hathitrust.org/cgi/pt?id=mdp.39015008971890&view=1up&seq=7> [2019. július 1.]
 13. MIKSA, Francis: The development of classification at the Library of Congress. Champaign (IL), University of Illinois Graduate School of Library and Information Sci-

- ence, 1984. (Occasional Papers 164.) Forrás: <https://www.ideals.illinois.edu/bitstream/handle/2142/3957/gslisoccasionalpv00000i00164.pdf> [2019. július 10.]
14. BLISS, Henry E.: A bibliographic classification, extended by systematic auxiliary schedules for composite specification. (4 köt.) New York, H. W. Wilson, 1940–1953.
 15. SEC. 901.” The National Science Foundation shall establish a Science Information Service. The Foundation, through such Service, shall (1) provide, or arrange for the provision of, indexing, abstracting, translating, and other services leading to a more effective dissemination of scientific information, and (2) undertake programs to develop new or improved methods, including mechanized systems, for making scientific information available.” Forrás: <https://www.govinfo.gov/content/pkg/STATUTE-72/pdf/STATUTE-72-Pg1580.pdf> [2019. július 10.]
 16. FUTALA Tibor: A VINITI ötven éve. = Tudományos és Műszaki Tájékoztatás, 50. évf. 6–7. sz. 2003. 291. p. A tömörítvény forrásműve: GILÁREVSKIJ, R. S. – CERNYJ, A. I.: Pát’desát let VINITI. Cto dal’s’e? Istoriá, dostizenia, budusee. = Naucno-tehniceskaá informáciá, 1 ser.; 2 ser. 12. no. 2002. 1–14. p. Ma Összorosországi Tudományos és Műszaki Információs Intézet néven működik, rövidítése maradt VINITI.
 17. A magyarul is megjelent változat: SHANNON, Claude E. – WEAVER, Warren: A kommunikáció matematikai elmélete. Az információelmélet születése és távlatai. Budapest, OMIKK, 1986.
 18. „[...]1950, Calvin Mooers wrote „The problem under discussion here is machine searching and retrieval of information from storage according to a specification by subject... It should not be necessary to dwell upon the importance of information retrieval before a scientific group such as this for all of us have known frustration from the operation of our libraries – all libraries, without exception.” MOOERS, C. N.: The theory of digital handling of non-numerical information and its implications to machine economics. In: Association for Computing Machinery Conference, Rutgers University, 1950. Idézi: SANDERSON, Mark – CROFT, W. Bruce: The history of information retrieval research. In: Proceedings of the IEEE. 100. vol. 2012. 1444–1451. p. Forrás: <https://dx.doi.org/10.1109/JPROC.2012.2189916> [2019. július 10.]
 19. TAUBE, M. et al.: Unit Terms in coordinate indexing. = American Documentation. 3. vol. 4. no. 1952. 213–218. p. Forrás: <https://doi.org/10.1002/asi.5090030404> [2019. július 1.]
 20. LUHN, H. P.: Keyword-in-context index for technical literature. In: Chan, L.M. et al. (ed.): Theory of subject analysis. A sourcebook. Littleton (CO), Libraries Unlimited, 1985. 168–180. p.
 21. LUHN, H. P.: The automatic creation of literature abstracts. = IBM Journal of Research and Development, 2. vol. 2. no. 1958. 159–165. p. Forrás: <https://doi.org/10.1147/rd.22.0159> [2019. július 10.]
 22. LERCH, Irmgard: Das Dokumentationshilfsmittel Thesaurus Darstellung der Entwicklung an Hand ausgewählter Beispiele. = Bibliothek – Forschung und Praxis, 6. Jg. 1–2. H. 1982. 47–73. p. Forrás: <https://doi.org/10.1515/bfup.1982.6.1.47> [2019. július 10.]
 23. VICKERY, B.C.: Thesaurus – a new word in documentation. = Journal of Documentation, 16. vol. 4. no. 1960. 181–189. p. Forrás: <https://doi.org/10.1108/eb026286> [2019. július 1.]
 24. Vö. „[...] Sparck Jones: „An Information retrieval thesaurus is a familiär object: it consists of a set of more or less controlled terms functioning äs conceptual labels for sets of

- natural language entry words, with some indication of relations between terms in the form of references from terms to other BT's, NT's or RT's². Such an indexing language differs from a set of natural language keywords in involving vocabulary normalisation; from a list of subject headings in being designed for post-co-ordination; and from a classification in having a limited structure." Lerch (1982) 47. p.
25. „This paper considers the possible utility of a citation index that offers a new approach to subject control of the literature of science." GARFIELD, Eugene: Citation indexes for science. A new dimension in documentation through association of ideas. = *Science*. 112. vol. 1955. 108. p. Forrás: <https://doi.org/10.1126/science.122.3159.108> [2019. július 10.]
 26. SALTON, Gerard: Automatic information organization and retrieval. New York, McGraw-Hill, 1968.
 27. SALTON, Gerard (ed.): The SMART retrieval system. Experiments in automatic document processing. Englewood Cliffs (NJ), Prentice-Hall, 1971. – A klaszterelemzésen nyugvó keresésről 223–242. p.
 28. SPARCK Jones, K.: A statistical interpretation of term specificity and its application in retrieval. = *Journal of Documentation*, 28. vol. 1. no. 1972. 11–21. p. Forrás: <https://dx.doi.org/10.1.1.115.8343&rep=rep1&type=pdf> [2019. július 10.]
 29. BROUGHTON, Vanda: The need for a faceted classification as the basis of all methods of information retrieval. = *Aslib Proceedings. New Information Perspectives*. 58. évf. 1–2. sz. 2006. 49–72. p. Forrás: <http://dx.doi.org/10.1108/00012530610648671> [2019. július 10.]
 30. BROUGHTON, Vanda: Brian Vickery and the Classification Research Group. The legacy of faceted classification. 2011. http://www.iskouk.org/sites/default/files/ISKOUK-2011_VandaBroughton.pdf [2019. július 10.]
 31. Vickery általános fazettáinak egyik változata: Substance (product), Organ, Constituent, Structure, Shape, Property, Object of action (patient, raw material), Action, Operation, Process Agent, Space, Time. VICKERY, B. C.: Faceted classification. A guide to the construction and use of special schemes. London, ASLIB, 1960.
 32. SHAPIRO, F. R.: Coinage of term Information Science. = *Journal of American Society for Information Science*, 46. vol. 5. no. 1995. 384–385. p. Forrás: [https://doi.org/10.1002/\(SICI\)1097-4571\(199506\)46:5<384::AID-ASI8>3.0.CO;2-3](https://doi.org/10.1002/(SICI)1097-4571(199506)46:5<384::AID-ASI8>3.0.CO;2-3) [2019. július 1.]
 33. FOSKETT, A. C.: The subject approach to information. 2. rev. ed. [s.l.], Linnet Books & Clive Bongley, 1971. 74–74. p.
 34. GARDIN, J. C. et al.: Le Syntol. Étude d'un système général de documentation automatique. Bruxelles, Presses Académiques Européennes. 1964. 4 köt. GARDIN, J.-C.: Recherches sur l'indexation automatique des documents scientifiques. = *Revue d'Informatique et de Recherche Opérationnelle*, 1. vol. 6. no. 1967. 27–46. p.
 35. AUSTIN, Derek: PRECIS. A manual of concept analysis and subject indexing. London, Council of the British National Bibliography, 1974.
 36. CLEVERDON, Cyril: Report on the testing and analysis of an investigation into the comparative efficiency of indexing systems. Cranfield, The College of Aeronautics. 1960. Forrás: <https://doi.org/10.1108/eb049778> [2019. július 10.]
 37. A Bliss Classification Association nevű társaság 1977-ben kezdte megjelentetni a második kiadást. Forrás: <http://www.blissclassification.org.uk> [2019. július 10.]

38. OHLY, H. Peter: Ingetraut Dahlberg. In: Encyclopedia of knowledge organization. ISKO, 2018. Forrás: <https://www.isko.org/cyclo/dahlberg> [2019. július 10.]
39. DAHLBERG, Ingetraut: Grundlagen universaler Wissensordnung. Pullach, Verlag Dokumentation, 1974. – A doktori disszertációja könyv alakban. Forrás: <https://doi.org/10.1515/9783111412672> [2019. július 1.]
40. Deutsches Institut für Normung. DIN 1463 1972; International Organization for Standardization ISO 2788-1974; American National Standards Institute. ANSI Z39.19-1974
41. HJØRLAND, Birger: i. m. 2008. 93. p.
42. MISHLER, Brent D.: History and theory in the development of phylogenetics in botany. In: Hamilton, A. (ed.): The evolution of phylogenetic systematics. University of California Press, 2014. 189-210. p. [http://ib.berkeley.edu/courses/ib200/readings/Mishler%202014%20\(History\).pdf](http://ib.berkeley.edu/courses/ib200/readings/Mishler%202014%20(History).pdf) [2019. július 10.] MISHLER, B. D.: History & philosophy of phylogenetics. Lecture. 2018. Forrás: <http://ib.berkeley.edu/courses/ib200/lect/lect03.pdf> [2019. július 10.]

Csik Tibor könyvtáros és történész, az ELTE BTK Könyvtár- és Információtudományi Intézet adjunktusa, a Könyv és Nevelés folyóirat főszerkesztője. Kutatási területe az ismeret-szervezés, valamint a könyvtár- és információtudomány története.

ORCID: 0000-0003-2061-1506